FIS Data Model Approach
Overview of SEFSC Using Dimensional Modeling

Overview of Dimensional Modeling

Benefits of the Approach

Tips and Techniques we Discovered Along the Way

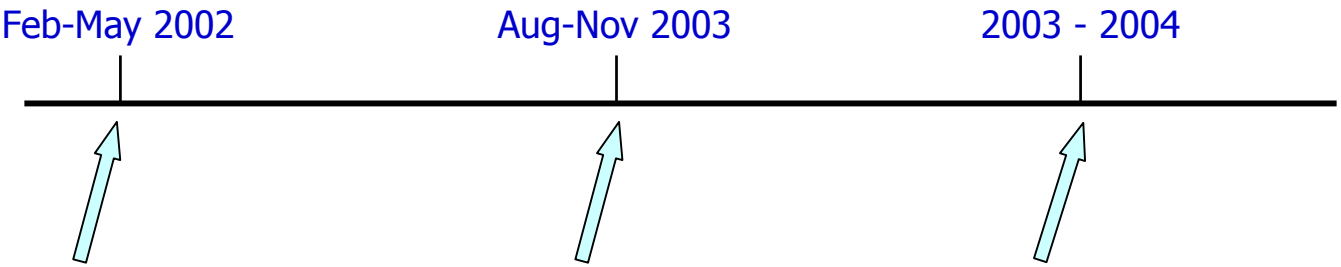
Agenda

- High Level Project Overview
  - Project Phases and Status
  - Role of Metadata and Oracle Warehouse Builder (OWB)
  - Goals
    - Goals of the Data Warehouse Project
    - Goals of the Extract-Transform-Load (ETL) Phase
  - ETL Task Diagram
- Dimensional Modeling
  - Facts, Measures and Dimensions
- NOAA ETL Architecture
  - ETL Process Flow

Presentation Does Not Focusing On…

- Types of tools and brand names
- Project management practices
- Details about Metadata use

SEFSC Data Warehouse Project Phases



Phase I
Requirements and Design
- ***Gather requirements***
- ***Design warehouse***
- ***Build proof of concept***
- ***Present prototype***

Phase II
Extract-Transform-Load
- ***Review prior work***
- ***Dimensional Modeling***
- ***Build Staging***
- ***Load Data Presentation Area***

Phase III
Data Access
- ***Configure access with OLAP tools***
- ***Setup security***
- ***Build report templates***
- ***Training and rollout***

Metadata

Phase I:	Design and prototype (proof of concept)
Phase II:	Extract-Transform-Load
Phase III:	Data access and presentation
Parallel Effort:	Metadata

Role of Metadata and OWB

- Metadata is all the information in the warehouse that is not the actual data
- Source schemas and access views
- Staging processes, transformations, and cleansing rules
- ETL schedules, security and confidentiality settings
- OWB repository mappings, objects and tools

Metadata is akin to an encyclopedia for the data warehouse – additional Metadata is a separate and ongoing effort

Goals of the Data Warehouse

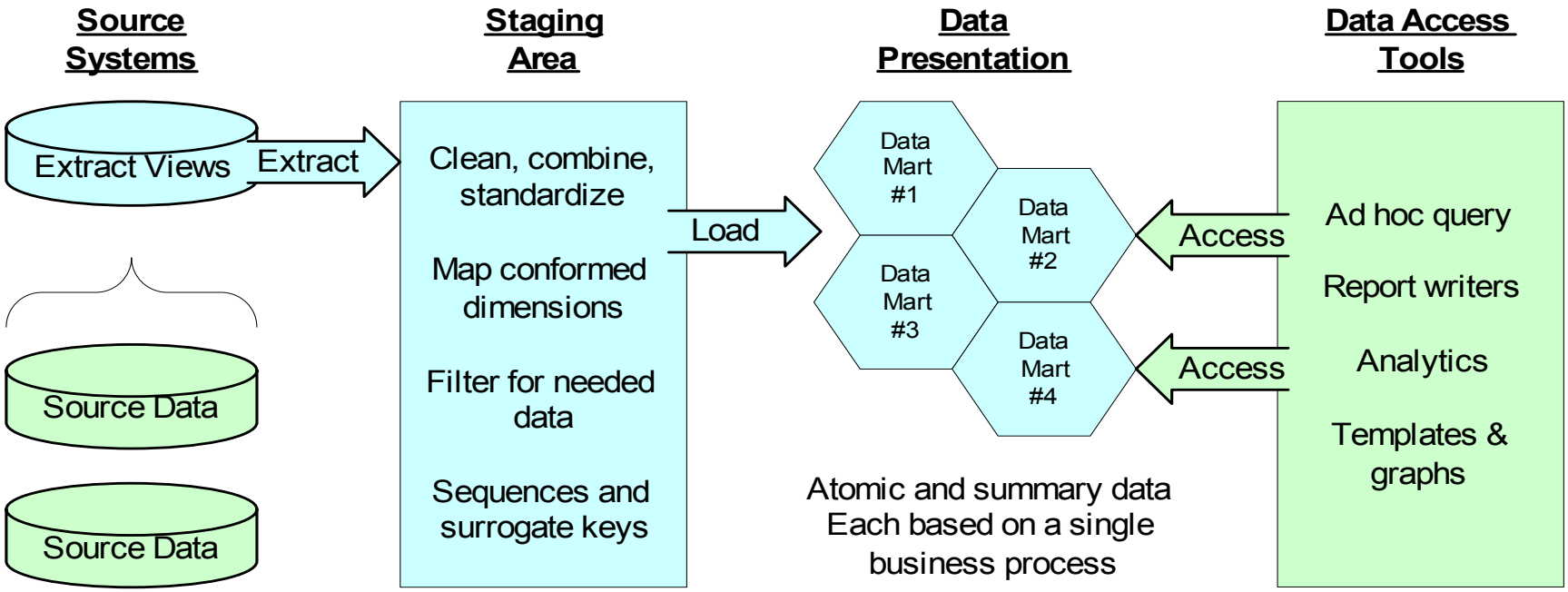
- Make information easily accessible
  - Content must be understandable
- Present information consistently
  - Data in the warehouse must be credible
- Must be adaptive and resilient to change
  - Design to handle changes gracefully
- Must be a secure repository
  - Maintain confidentiality
- Support improved decision making
  - Reliable source of analysis data for decisions

Goals of the ETL Phase

- Build extracts from the source systems
- Build the data staging area

  (between source systems and data presentation area)
- Clean, combine and standardize the data
- Populate the data presentation area using dimensional modeling methods

ETL Task Diagram



ETL tasks move operational source data to dimensional models

Agenda

- High Level Project Overview
  - ✓ Project Phases and Status
  - ✓ Role of Metadata and Oracle Warehouse Builder (OWB)
  - ✓ Goals
    - Goals of the Data Warehouse Project
    - Goals of the Extract-Transform-Load (ETL) Phase
  - ✓ ETL Task Diagram
- Dimensional Modeling
  - Facts, Measures and Dimensions
- NOAA ETL Architecture
  - ETL Process Flow

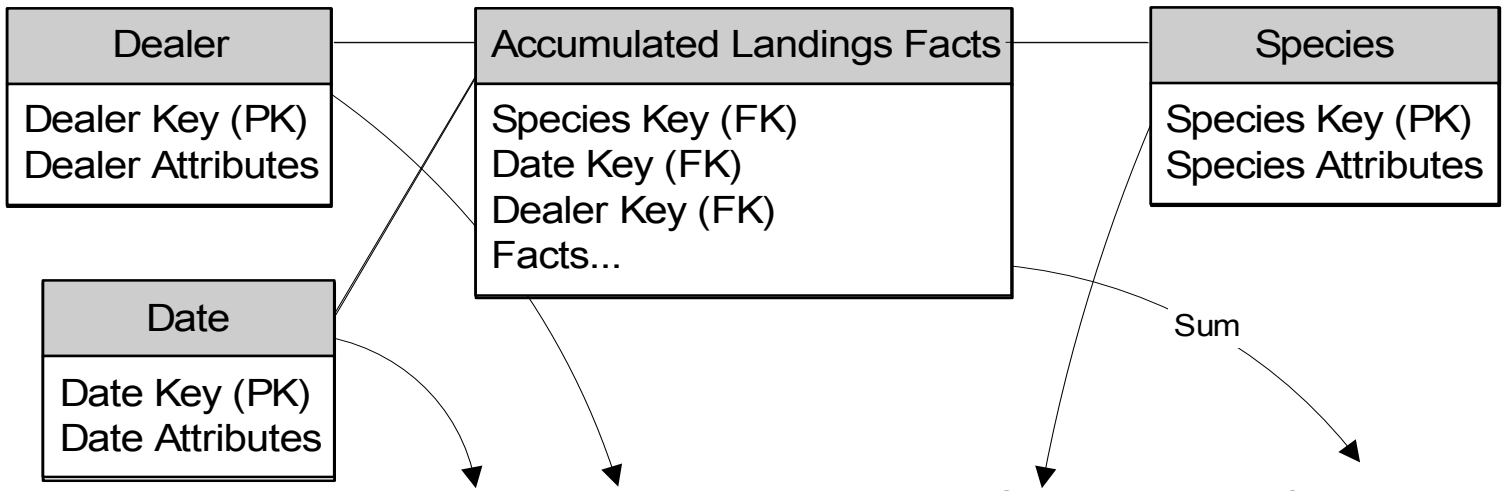
Elements of Dimensional Modeling

- Facts (measurements)
  - The primary table to a dimensional model
  - A row corresponds to a measurement
  - Best if numeric and additive
  - Measurement data from a single business process stored in single data mart
  - Measurements are the intersection of dimensions
- Example
  - Pounds Landed
  - Number of Samples taken

Elements of Dimensional Modeling

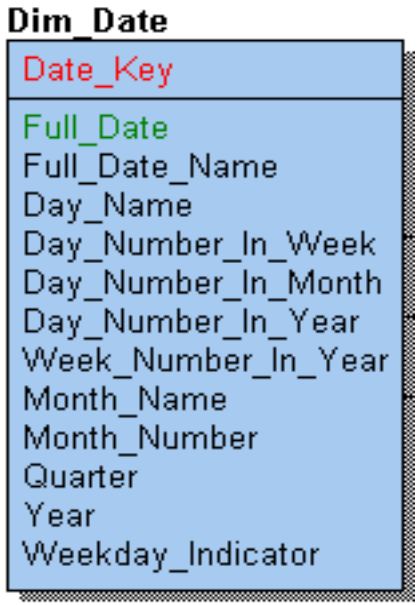
- Dimensions
  - Textual description of the business
  - The *by* words
    - See the measure by month, by species, by dealer
  - Query constrains, groupings, report labels
  - The entry point into the Fact tables
  - Robust dimensions attributes provide dicing and slicing analytic capability
- Example
  - Species
  - Dealers

Dimensional Model (example)



Date	Dealer	Species	Quantity Landed (lbs)
January	Harrells Fish Market	Blue Catfish	59.4
January	Harrells Fish Market	Flathead Catfish	14.0
January	Harrells Fish Market	Striped Mudfish	89.0
January	McCabes Grocery	Blue Catfish	35.0
January	McCabes Grocery	Flathead Catfish	57.0
January	McCabes Grocery	Striped Mudfish	29.0

The Date Dimension (a closer look)



- Built in advance
  - Range 1960 – 2010
- Surrogate key (red)
- Natural key (green)
- Conformed dimension (blue)
- May provide Date attributes not supported by SQL Date function
  - E.g. Lobster season
- Change/expand gracefully
  - E.g. Decade attribute

The Date Dimension (a closer look)

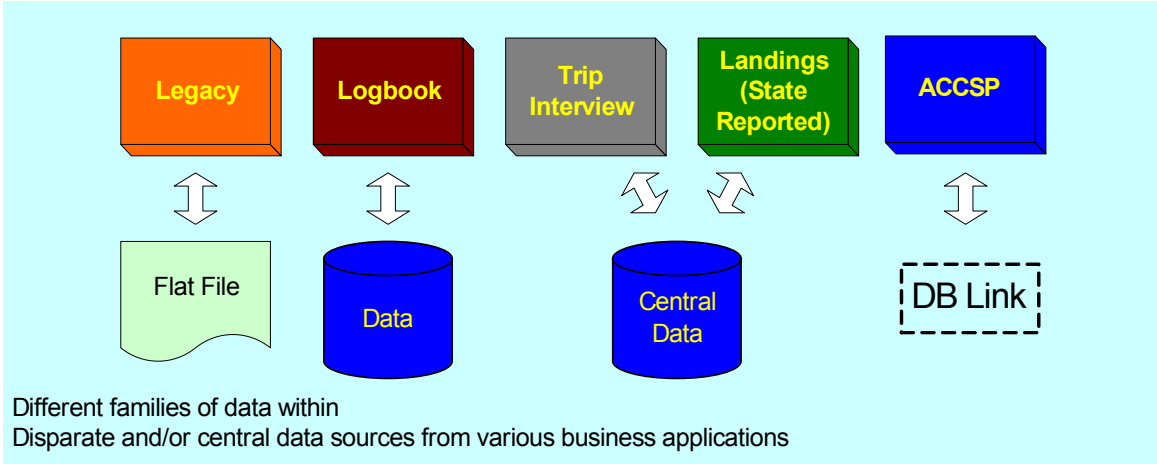
DAY_KEY	FULL_DATE	FULL_DATE_NAME	DAY_NAME	WEEK IN YR	MONTH_NAME	QUARTER	YEAR	WEEKDAY_INDICATOR
18641	1/1/1960	January 1, 1960	FRIDAY	1	JANUARY	1	1960	WEEKDAY
18642	1/2/1960	January 2, 1960	SATURDAY	1	JANUARY	1	1960	WEEKEND
18643	1/3/1960	January 3, 1960	SUNDAY	1	JANUARY	1	1960	WEEKEND
18644	1/4/1960	January 4, 1960	MONDAY	1	JANUARY	1	1960	WEEKDAY
18645	1/5/1960	January 5, 1960	TUESDAY	1	JANUARY	1	1960	WEEKDAY
18646	1/6/1960	January 6, 1960	WEDNESDAY	1	JANUARY	1	1960	WEEKDAY
18647	1/7/1960	January 7, 1960	THURSDAY	1	JANUARY	1	1960	WEEKDAY
18648	1/8/1960	January 8, 1960	FRIDAY	2	JANUARY	1	1960	WEEKDAY
18649	1/9/1960	January 9, 1960	SATURDAY	2	JANUARY	1	1960	WEEKEND
18650	1/10/1960	January 10, 1960	SUNDAY	2	JANUARY	1	1960	WEEKEND
18651	1/11/1960	January 11, 1960	MONDAY	2	JANUARY	1	1960	WEEKDAY
18652	1/12/1960	January 12, 1960	TUESDAY	2	JANUARY	1	1960	WEEKDAY
18653	1/13/1960	January 13, 1960	WEDNESDAY	2	JANUARY	1	1960	WEEKDAY
18654	1/14/1960	January 14, 1960	THURSDAY	2	JANUARY	1	1960	WEEKDAY
18655	1/15/1960	January 15, 1960	FRIDAY	3	JANUARY	1	1960	WEEKDAY
18656	1/16/1960	January 16, 1960	SATURDAY	3	JANUARY	1	1960	WEEKEND
18657	1/17/1960	January 17, 1960	SUNDAY	3	JANUARY	1	1960	WEEKEND
18658	1/18/1960	January 18, 1960	MONDAY	3	JANUARY	1	1960	WEEKDAY
18659	1/19/1960	January 19, 1960	TUESDAY	3	JANUARY	1	1960	WEEKDAY

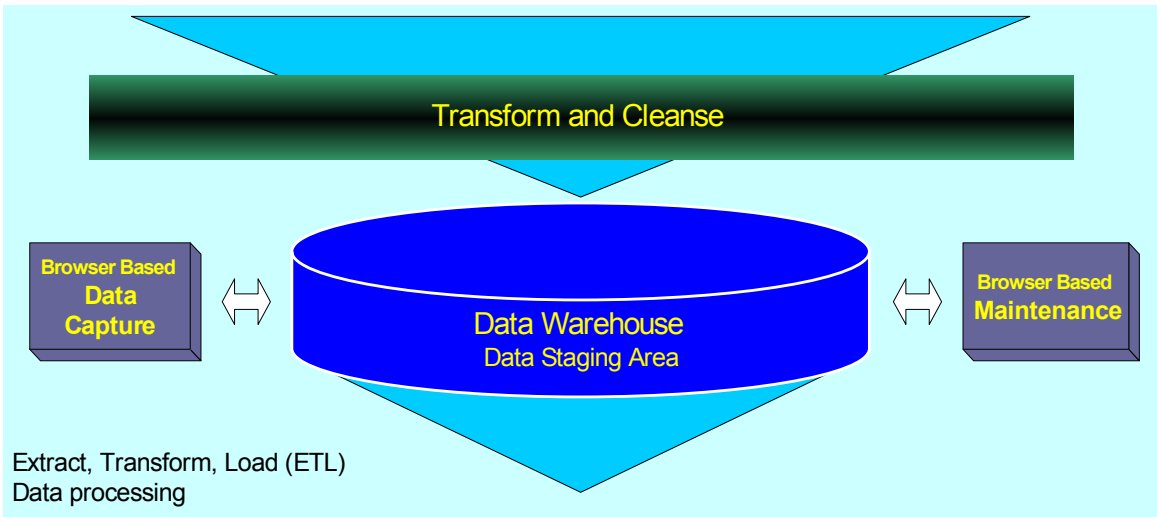
Elements of Dimensional Modeling

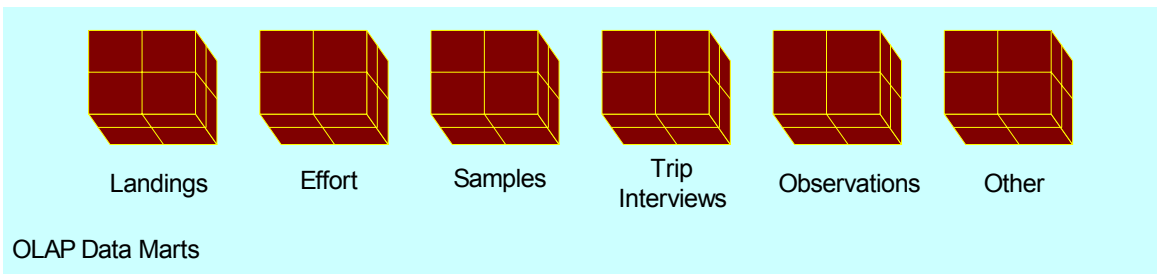
- Types of Fact table measures
  - Transaction
    - At the individual transaction level
    - Represent an event that occurred at an instantaneous point in time
  - Periodic snapshot
    - See cumulative performance at regular intervals
    - Show a trendable view of performance metrics
  - Accumulating snapshot
    - Indeterminate time span covering a complete life of a transaction/series of events
    - Multiple data stamps for predictable major events

Agenda

- High Level Project Overview
  - ✓ Project Phases and Status
  - ✓ Role of Metadata and Oracle Warehouse Builder (OWB)
  - ✓ Goals
    - Goals of the Data Warehouse Project
    - Goals of the Extract-Transform-Load (ETL) Phase
  - ✓ ETL Task Diagram
- Dimensional Modeling
  - ✓ Facts, Measures and Dimensions
- NOAA ETL Architecture
  - ETL Process Flow

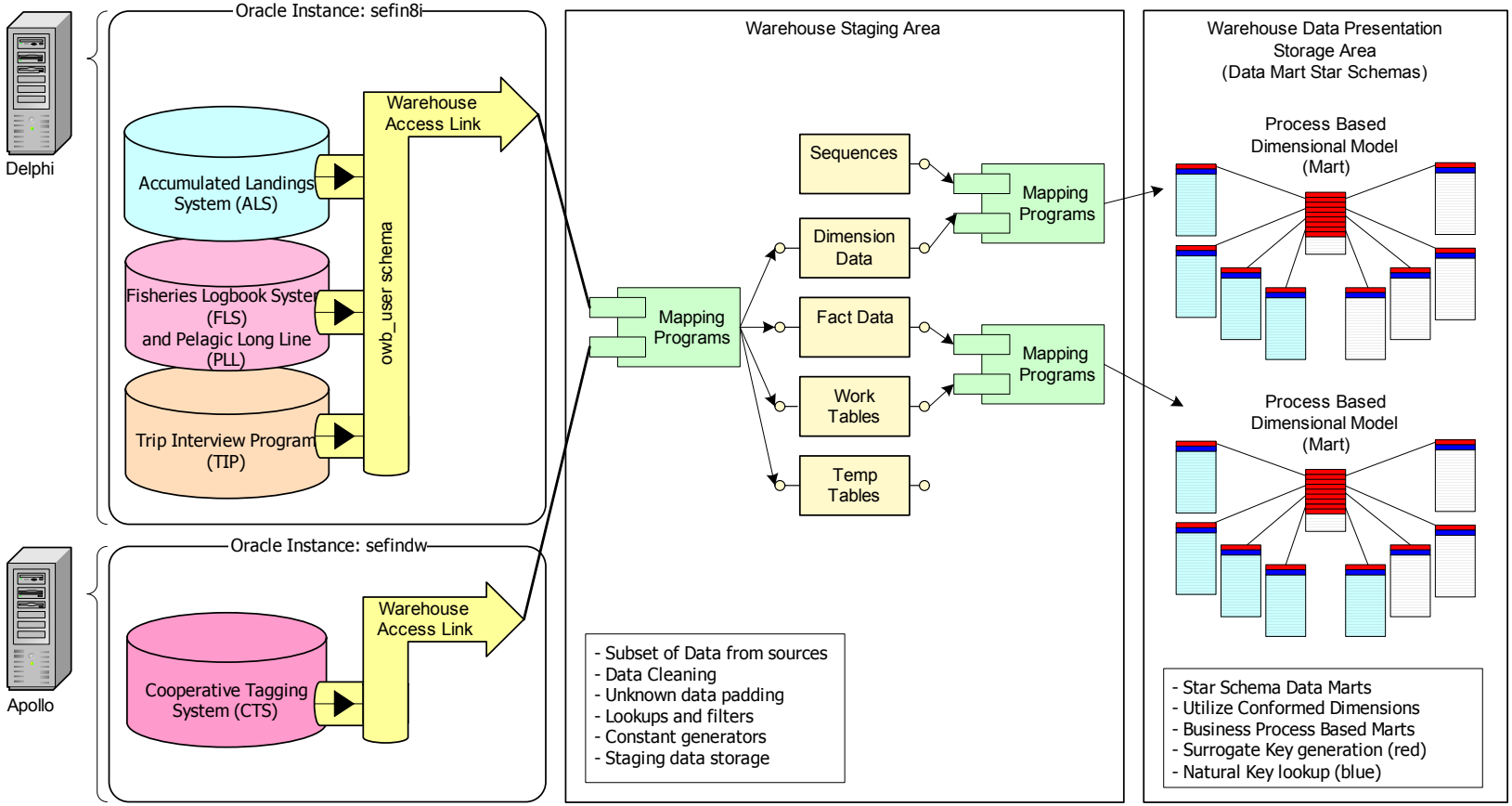




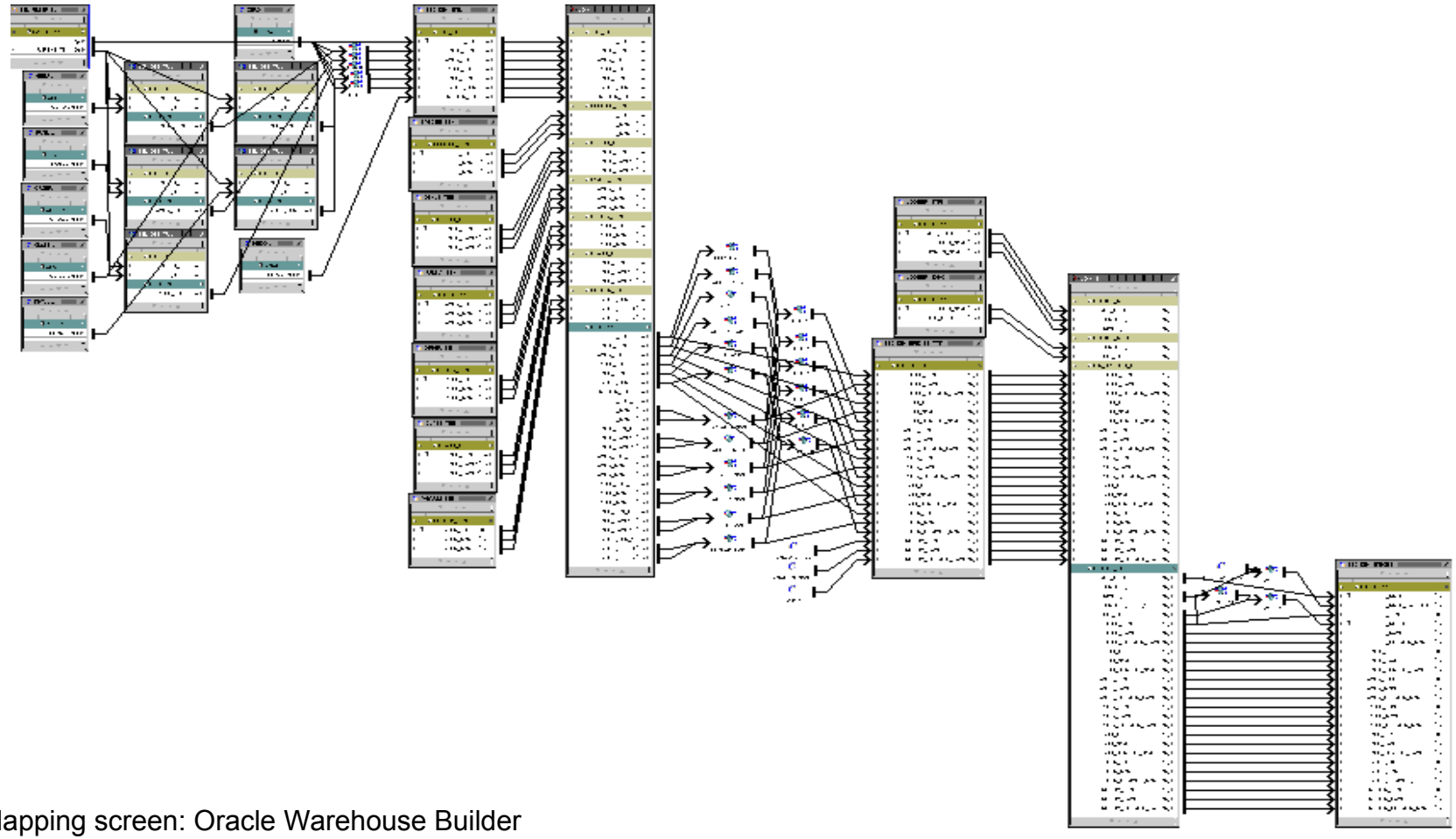


The multi-tier or "hub-and-spoke" data warehouse features a general-purpose relational data staging area as the "hub", coupled with OLAP-based application specific data marts as "spokes" to deliver information efficiently

SEFSC ETL Architecture

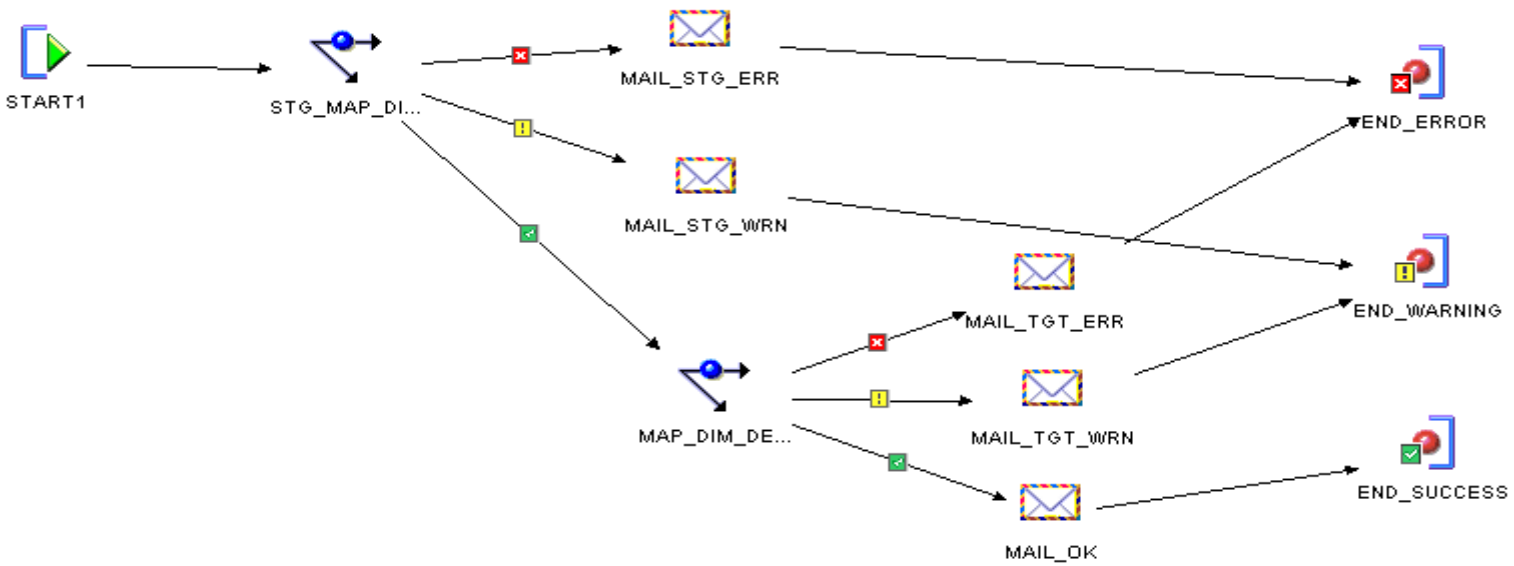


Example Mapping



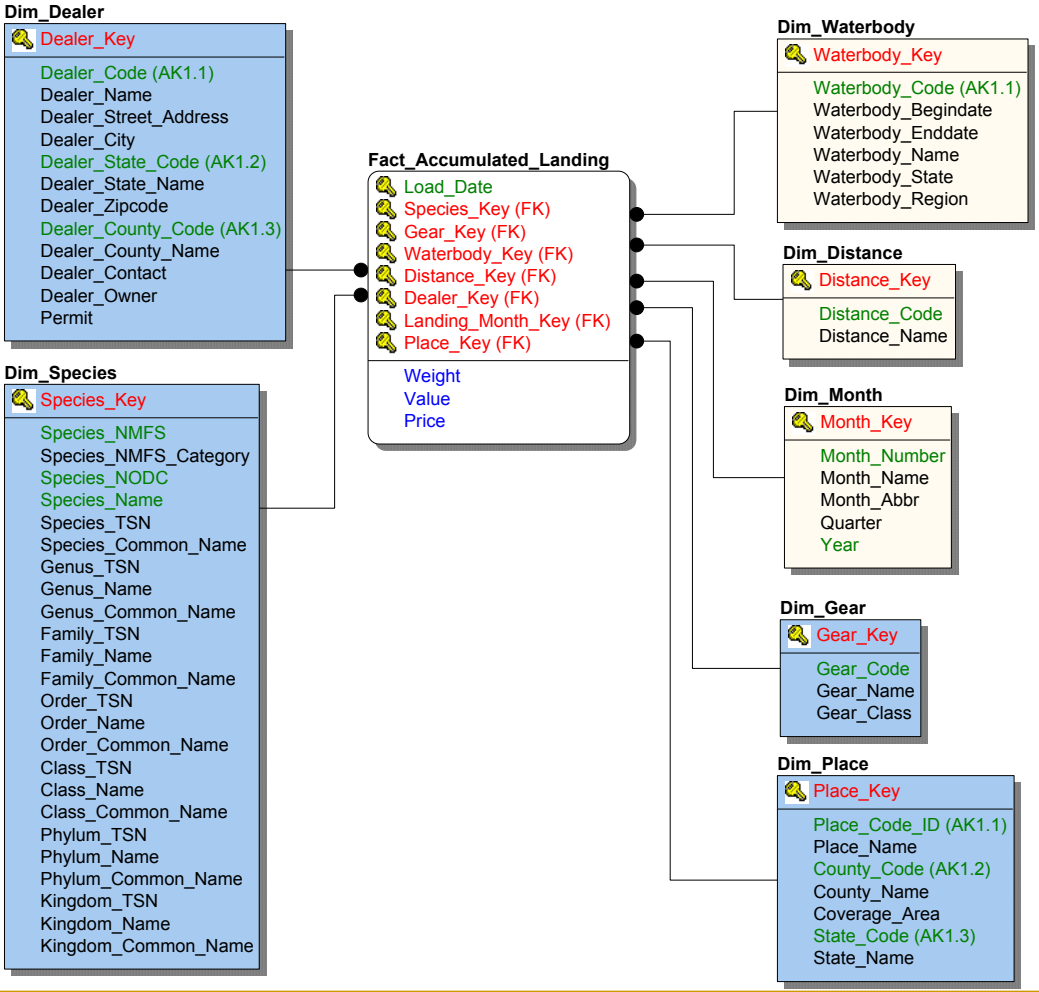
Mapping screen: Oracle Warehouse Builder

Example Work Flow

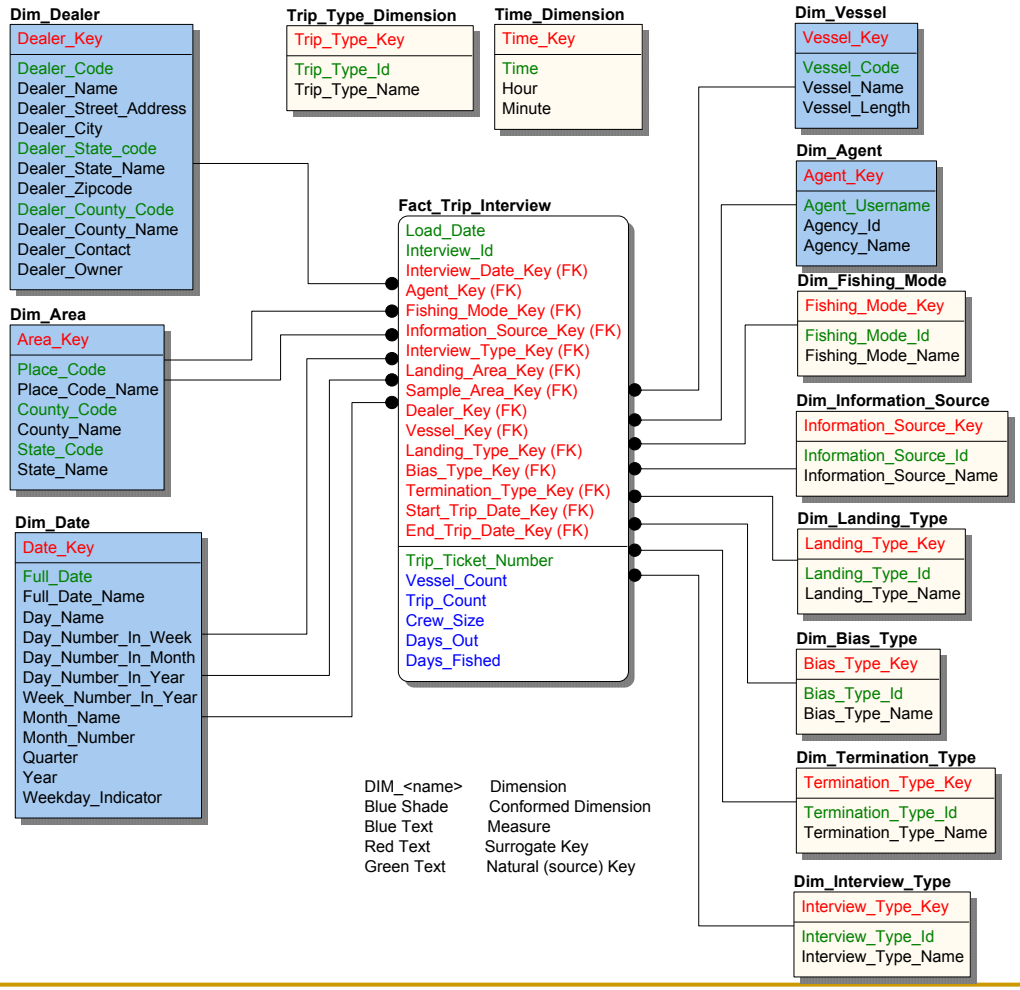


Workflow Design Screen: Oracle Warehouse Builder

Other models: Accumulated Landings



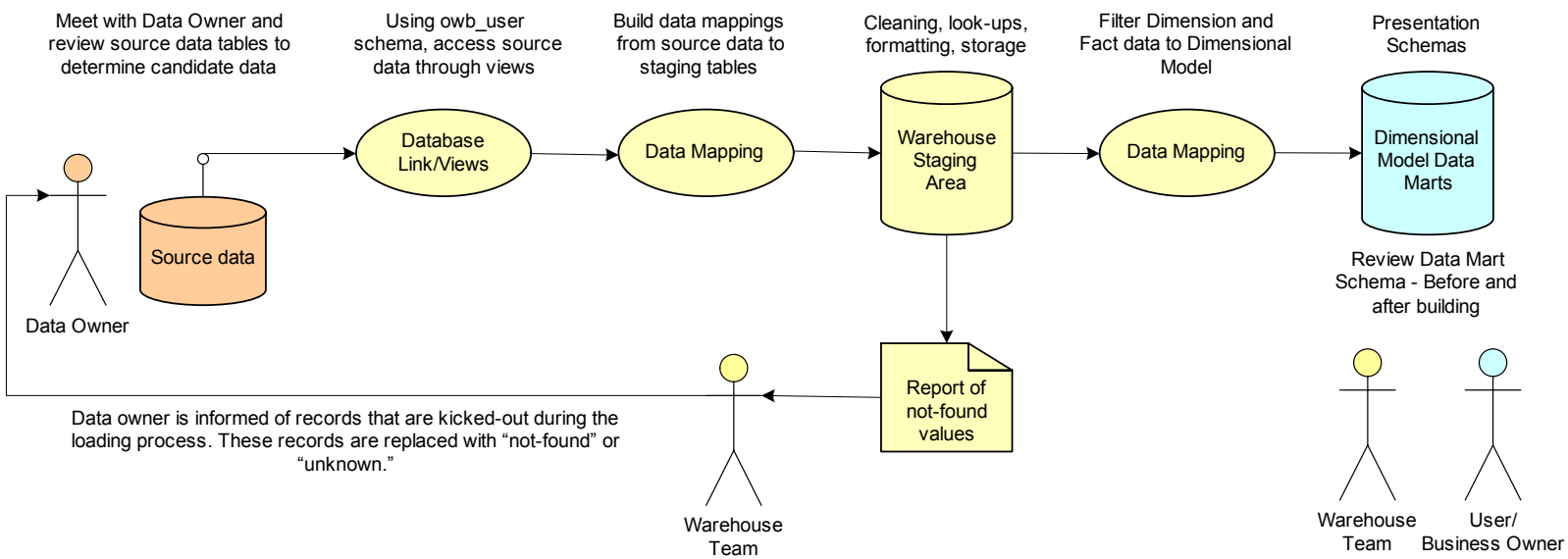
Trip Interview Model



Trip Interview Report

YEAR	1993												
Count of Sum (Days Fished)	MONTH_NAME												
VESSEL_NAME	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Grand Total
2ND DESTINY	1											1	2
2ND WIND	1							1			1	1	4
786 BENGAL I			1										1
ADELAIDE				1									1
ADVENTUROUS												1	1
AILEEN II	1								1			1	3
ALEX						1					1		2
ALEX JAMES					1								1
ALEXIS M	1								1			1	3
ALLANA KAY	1								1		1	1	4
ALMOST	1												1
AMY MARIE						1							1
AMY MICHELLE	1	1	1	1	1	1	1	1	1	1			10
ANGLER	1	1	1		1			1	1				6

ETL Workflow



Unknown & Orphan records

- Unknown records
  - Provide a value to represent 'Unknown'
  - Optional data that is missing becomes valued as Unknown
- Orphan
  - We have Fact data without Dimension Data
  - Generate Orphan records and report
  - Allow loading process to continue

Questions